THE GRADUATE INSTITUTE GENEVA

CENTRE FOR INTERNATIONAL ENVIRONMENTAL STUDIES

Research Paper 60 | **2020**

CLEAN ENERGY INNOVATION AND THE INFLUENCE OF VENTURE CAPITALISTS' SOCIAL CAPITAL

Till FUST

THE GRADUATE INSTITUTE GENEVA

CENTRE FOR INTERNATIONAL ENVIRONMENTAL STUDIES

Clean Energy Innovation and the Influence of Venture Capitalists' Social Capital

Till Fust[1]

September 2020

Abstract

This study contributes to the understanding of the enabling role that venture capitalists can play in bringing new innovative technologies to market, with a focus on clean energy technologies. Applying the structural model introduced by Sørensen (2007) that allows to control for a potential sorting bias, I estimate the influence of venture capital investor's social capital on startups' funding and exit performance, with social capital defined as the investors' eigencentrality and constraint within the network of investors. Looking at startups' first venture capital funding rounds in California between 2001 and 2019, this study finds a positive and significant influence of the lead investor's eigencentrality on the funding amount raised and the exit probability of the firm. Furthermore, a less constrained lead investor also increases the chance of the startup's eventual exit. But no differentiated effect for cleantech startups compared to other industries is found.

[1]This study is supported by the Swiss National Science Foundation (SNSF) within the framework of the National Research Programme "Sustainable Economy: resource-friendly, future-oriented, innovative" (NRP 73, Grant-No 407340-172395). I would like to thank Joëlle Noailly and James Hollway for their support and inputs. Corresponding author, Centre for International Environmental Studies, Graduate Institute of International and Development Studies. Email: till.fust@graduateinstitute.ch.

Contents

1 Introduction

1.1 Venture capital and innovation in clean energy

According to the IPCC (2018) report, limiting global warming to 1.5 or 2 degrees Celsius requires net zero emission by 2050 or 2070, respectively. This requires replacing great amounts of non-renewable energy sources driving our current economy and everyday lives and ”without a major acceleration in clean energy innovation, net-zero emissions targets will not be achievable” (International Energy Agency (2020)). Facing climate change, the question of how innovation in the clean energy technology (cleantech) sector can be fostered and supported has gained more and more traction. With respect to financing innovation, in recent years venture capital (VC) has begun to play an important role in supporting innovation across multiple sectors. When Gornall and Strebulaev conducted a study on the extent of the VC industry in the USA with data from 2014, they found that of all public companies, 17% were VC backed in their early stages, constituted 21% of the total market capitalisation and almost half (44%) of the R&D spending by public companies. When only looking at companies founded after VC was an available option to entrepreneurs, the numbers are even higher with 42% VC-backed public companies (Gornall and Strebulaev (2015)). These statistics are at least partly driven by very big successful outliers like Google, Apple or Microsoft, but it does not include private companies that were VC backed and shows the potential impact the VC industry can have on the economy. Venture capital investors (VCI) bundle high-risk, high-reward investment profile startups, meaning that a significant portion of the investments are at risk of failing but the higher rewards from a few successes should result in a positive expected return. This gives innovative startups the financial leverage to prove marketability of their novel concept when the risk of failure is high. For example, in the investment data from 1981 to 2005 that Puri and Zarutskie (2012) use for their analysis of VC investments, 47% of all firms receiving VC funding start without any commercial revenue.

Analysing the effect of VC on innovation, Kortum and Lerner (2000) find a strong positive effect of the increasing VC industry in the USA from 1983 to 1992. Faria and Barbosa (2014) find a positive effect of VC on patent applications for European firms for later stage VC investments and see evidence for the role of the VC firm rather in helping to commercialize innovation but not in inducing it. Furthermore, for European data, Geronikolaou and Papachristou (2012) find evidence for the causality running from innovation to VC and not the other way around. Similarly, Hirukawa and Ueda (2011) examine the hypothesis of whether it is VC that spurs innovation or whether innovation leads to more VC investments. They find evidence for the causality of the latter. Since an increasing number of young companies choose VC as their funding instrument, these findings indicate that even though VC might not be the inducer of innovation, it can at least be seen as a complement to innovation, acting as an enabler helping to bring innovations to market.

Looking at VC funding as an enabling instrument in cleantech, a sharp increase in VC funding into clean energy technology firms was seen after 2004, but this dropped again after 2011 with early stage funding even earlier in 2008, as the returns were not satisfying enough (Gaddy et al. (2016)). In an analysis of this downturn, Gaddy et al. (2016) suggest that VC may not be well suited for investments into cleantech because of a longer time horizon, higher capital intensity and the greater difficulty to exit. Two of the stated reasons for the investment boom into cleantech to drop again were the 2008 economic crisis as well as oil prices not increasing as predicted at the time of the boom, which led to low returns of the VC cleantech investments. Additional research points to the long time frame causing a higher risk of exposure to the "Valley of Death" which refers to the financial and timely gap between the basic research and the commercialization of an innovation (Nanda et al. (2014), Popp et al. (2020)). However, Popp et al. (2020) also see a changing energy market which has smaller and more modular technologies and a higher dependency on advancements in other sectors. They find that a growing percentage of high-tech energy startups are more likely to receive VC funding (but do not perform better than "normal" clean energy startups) and highlight that when looking at numbers of new patents for clean technologies, even though they decreased since 2010 alongside VC funding, the sharp increase in 2004 to 2011 might just have been a bubble. Compared to 2006, patent count in 2015 was 16% higher which is a growth rate above the average of other industries. Popp et al. (2020) stress the changing innovation in clean energy and the higher complexity that comes with it. They additionally find descriptive evidence that energy startups potentially attract higher funding, especially those with high-tech components. The recently released report by the International Energy Agency (2020) with data up until 2019 now shows an increase to even above 2010/2011 levels in seed, series A and series B venture capital funding of cleantech startups in recent years. Even hardware technologies that are more prone to the "Valley of Death" (Gaddy et al. (2016)) are seeing an increase in funding.

Overall, existing literature suggests that in the past decades, VC has complemented innovation in various industries and helped bringing new technologies to market. In the case of cleantech, after a boom between 2004 to 2011, the confidence in the marketability of clean energy technologies was negatively impacted due to various circumstances and therefore missing returns. This led to the reasoning that VC might not be a suitable financing tool for the sector. However, recent findings suggest that the 2004 to 2011 period might have just been a boom followed by sharp decline. The latest data now even report a strong incline in clean energy VC investments, showing that investor confidence might have at least partly rebounded.

1.2 Channels for venture capitalists' influence

Given that VC funding as an instrument has helped various innovations reach marketability in the past and that latest research suggests that there might be a place for VCIs in the funding of clean energy technologies, this research will seek to answer the following question: Through which channels can VCIs impact the performance of their investments and thus act as enablers for innovation?

One mechanism highlighted above is the risk structure of VCIs' portfolios, giving firms developing unproven innovations the time needed to build their proof of concept. But VC investments usually do not only come with financial resources, but the VCIs often also take on a guidance or mentorship role for the young firm, contributing to their investments' success directly with their own knowledge and indirectly by providing access to their network and acting as a resource and information broker. Eventually, the VCIs' success depends on the ability of their portfolio companies to create enough value to be acquired or to go public especially given young companies often lack stable management or networking resources. Some of the main channels of influence are the provision of financial and strategic business advice, the interpersonal relationship as mentor, reputation gains (signaling), discipline through monitoring and access to the investors own networks (Sapienza et al. (1996), De Clercq et al. (2006)). The findings of Hsu (2006) suggest that VC-backed startups are also more likely to engage in cooperative activity. Using one-to-one matching of VC-backed with non-VC-backed firms, Puri and Zarutskie (2012) find that VC-backed firms grow faster and larger in scale in terms of employment and sales, which suggests that VCIs invest in firms with a high scalability. They also show that profitability at exit is not larger, but the exit rate is higher compared to firms not backed by VC. Paglia and Harjoto (2014), after controlling for the access to such funding, found that VC funding had a positive impact on the companies' net sales and employment growth. Chemmanur et al. (2011) report a higher total factor productivity growth in the early stages for VC-backed firms mainly due to increased sales and for high-reputation VCIs also trough lower production cost increases. They show that this effect is partly due to screening (VCIs screening the market for - and investing in - more productive enterprises), as well as the investors monitoring post-investment.

The main focus in this study is on one of these channels - the influence of the VCIs' networks. More specifically, the VCIs' social capital from their networks of investors that evolves from investment syndication.

1.3 Influence through social capital and research questions

The term "social capital" has been widely used in research and been defined in different ways in different contexts. According to Burt (2005), "...but they agree on a social capital metaphor in which social structure defines a kind of capital that can create for individuals or groups an advantage in pursuing their ends." An investor would likely build social capital from various social structures like friends, family, former jobs, school and more general contacts, but in the context of this research I follow previous literature (Alexy et al. (2012), Hochberg et al. (2007), Sorenson and Stuart (2001)) and abstract and define social capital as a resource that investors have (or not) through their ties and their position in the network of investors which evolves from investment syndication (the precise construction of this network is explained subsequently in section 3). Having a higher social capital can provide the investor with novel information that can be used to find investment opportunities, but also to share with the startup for their advantage. Better connected VCIs could facilitate further funding or an eventual exit of the company. Previous research on this has shown that VC funds led by high social capital VCIs have a significantly better fund performance and their portfolio companies are more likely to survive to further funding rounds and eventual exit (Hochberg et al. (2007)). Alexy et al. (2012) find that higher social capital is related to a higher funding amount in the first funding round of the company and Hsu (2004) shows that high-reputation VCIs have a higher chance that their financing offer is being accepted by the startup.

Existing literature indicates that startups financed by VCIs with a higher social capital do perform better, which speaks for the performance of well-connected VCIs and shows the value added for the investors to build social capital. However, there is limited research that disentangles the two effects that cause the better performance:

sorting and influence. The social capital from investor networks might increase the VCIs' portfolio performance simply through better screening of the investment market or a higher chance to invest in more promising deals and therefore investing in initially already superior firms. Regarding the cleantech perspective, the screening function of investors can be helpful in sorting the market into clean technologies that are worth funding and those that are not. But the main objective of this study is to analyse how the social capital of VCIs within their network of investors can influence the funding and the success of the startup post-funding and thus enable the innovation's eventual commercialisation.

I do so by applying a structural matching model introduced by Sørensen (2007). While he applies the model to analyse the two effects for the investors' experience and its effects on the investees' IPO probability, this study leverages it to evaluate the influence the VCIs' social capital has on startup funding and performance with an additional focus on cleantech. I test if VCIs with higher social capital are willing to pay a funding premium because before the initial investment they might have superior information about the investment environment and new ventures and therefore a lower risk, which allows them to pay that premium, and if post-investment they facilitate further funding rounds and a successful exit because they provide higher non-financial resources to help the startup succeed. While this might be the case for any VC investment, given the nature of the cleantech sector discussed before (longer time horizon, high capital intensity, potentially higher complexity) I argue that it could be even more important in this sector (better screening pre-investment and better guidance/help post-investment) to help overcome these challenges. I therefore test whether the influence is different for the cleantech sector. Answering these questions could help for a better understanding of the role VCIs can play as enabler for cleantech innovation.

The outline of this study is structured as follows. Section 2 justifies the choice of methodology and introduces the applied structural model by Sørensen (2007); in section 3 follows a description of the data and the choice and construction of the variables is explained; section 4 shows results of the baseline regressions only controlling for sorting on observed characteristics; section 5 then presents the main results from the structural model, section 6 discusses the results and section 7 concludes.

2 Methodology

To tackle the posed questions, I start by estimating the overall effect of social capital on the amount raised in the initial funding round (ols), the probability that further funding rounds occurred (probit), the total amount raised in the funding rounds following the initial round conditional on at least one additional funding round having occured (ols) and eventually the exit probability of the company (probit). These models give a first overview of the sorting bias on observed characteristics and provide a baseline for comparison and discussion of the results from the structural model, where the outcome equation is specified in the same way and the selection equation controls for sorting on unobserved characteristics.

2.1 Sorting and Influence - Structural Model

When evaluating the impact of investor characteristics on funding round outcomes and startup performance, an important distinction between two effects has to be made – sorting and influence. An investor's social capital, the investor characteristics of interest in this study, can potentially positively influence the startups performance through their access to helpful resources from within the investor's network. But the investor seeking a financial return will try to leverage this same network not only to help his investments succeed, but also to get the best deals initially. Investors with a higher social capital are likely to have a better reputation, thus being more attractive to startup when they know of the benefits of a reputable and experienced investor. When investors with higher social capital get access to better deals, then their investments are also likely to show higher funding amounts and have a higher chance of succeeding, which leads to biased estimates when this is not controlled for. Since in this work I focus on the influence rather than the success of investors' portfolios, this has to be accounted for.

One way to do so is to control for certain indicators of investor quality, like experience or past successful investments, to at least partly lower that bias. But there is only little information available to control for the quality of the startup. E.g. Alexy et al. (2012) do so by including a measure of serial entrepreneurship (whether the founders have previously founded other companies) as well as patent and trademark data. Other controls could potentially be grades and scores from startup competitions or the founder's education. This can give an approximation of the quality of the deal, but the chances of an omitted variables bias remain high, as characteristics like the quality of the management team can hardly be observed. Another common statistical approach would be to find an instrumental variable. Unfortunately, it seems nearly impossible to find a plausibly exogenous IV that has some explanatory power for the social capital and that can be argued to be independent of the initial investment amount or the success of the startup.

In this study I differentiate between sorting and influence by applying a structural model introduced by Sørensen (2007), suggesting leveraging exactly this underlying matching game in the startup-investor matching market. The most promising startups have better choices in terms of who they can take on as their investors and vice-versa more reputable investors are likely to have a better shot at investing in more promising startups. This leads to competition on the best possible matches on both sides, the investors as well as the startups, and eventually, the relative qualities of the investors and companies in each investment market decide the outcome of the matching game. The investments, or investor-startup matches, are the observed outcome of this underlying two-sided matching game. But in a given investment market, the startups could have matched with other investors and these unobserved potential matches represent non-equilibrium matches. Intuitively, we use the exogenous variation in the characteristics of the agents of unobserved, non-equilibrium matches, to impose restrictions on the latent match valuation that identify the structural model. The model is a special application of the college admission model introduced by Gale and Shapley (1962) and I will only summarise the basic setup of the model by Sørensen (2007), that is outlined in more detail in the original paper.

The outcome of the underlying matching game in each investment market, the observed matches, is said to be stable when no agent prefers to deviate from their current match. A stable matching always exists (Gale and Shapley (1962)). On the basis of the distinction between observed and unobserved matches we can now put restrictions on the latent match valuations in the estimation. Every unmatched pair (unobserved investments) of investor i and company j has an opportunity cost $\bar{V}_{ij}$ of deviating from the current match and form a new one together, which is essentially the valuation of company j's current match and the lowest valued investment in investor i's current portfolio in the market. When the match valuation V_{ij} of an investor-company pair ij is higher than this opportunity cost of deviating from the current match, investor i and company j will prefer to leave their current match to form a new one together. When for all unmatched investor-company pairs the valuation is below $\bar{V}_{ij}$ no agents want to deviate, and the outcome is stable. Likewise, every matched pair (observed investments) has an opportunity cost $\underline{V}_{ij}$ of remaining together, which is the maximum match valuation, other than the current, that lies in a feasible set of matching partners. For investor i that is all organisation j' for which forming a new match with investor i would increase their match valuation, and likewise for organisation j all investors i' for which forming a new match with j and dropping the lowest valued matching pair in their portfolio would increase their total valuation. When V_{ij} is above $\underline{V}_{ij}$ for all observed matches, the outcome is stable. Given the observed matches μ, the set of valuation for which this represents a stable equilibrium is Γ_{μ}.

Empirically, the structural model consists of a selection and an outcome equation. The selection equation

$$V_{ij} = W'_{ij}\alpha + \eta_{ij}$$

with observed characteristics of investors and startups W_{ij} for all potential matches and independent error term η_{ij}. Since all latent valuations have to be in Γ_{μ} for the observed matches to represent a stable equilibrium, the matching is only stable if $\eta \in \Gamma_{\mu} - W\alpha$, which can be solved for by maximising

$$L(\mu, \alpha) = Pr(\eta \in \Gamma_{\mu} - W\alpha) = \int 1[\eta \in \Gamma_{\mu} - W\alpha]dF(\eta)$$

The second part of the structural model is the outcome equation

$$Y_{ij} = X_{ij}\beta + \epsilon_{ij}$$

with $\epsilon_{ij} = \delta\eta_{ij} + \zeta_{ij}$ where ζ_{ij} is a random error and δ measures the covariance between the error terms of the selection and the outcome equation. If δ was zero and insignificant, there would be no selection issue. If it is non-zero and significant, it captures unobserved factors that affect the selection as well as the outcome equation and eventually captures sorting over characteristics that are unobserved in the data.

Three important main assumptions of this model have to be discussed (Sørensen (2007)), which also drive some of the restrictions applied to the data analysed in this work.

First, the model is a one-to-many matching model. Each company can only match with one investor, while an investor can match with multiple startups. In reality, this is not the case. As we will see in the data for California, every initial funding round consists of 3 investors on average. This assumption restricts the sample for the analysis to the lead investors. Lerner (1994), in an analysis of VCI-syndication, finds that for first investment rounds, experienced VCIs co-invest with other investors with similar levels of experience, which at least partly justifies this assumption.

Second, investors can only invest a certain amount of money and time. Assuming the VCI's scarcer resource is time, every investor is assigned a certain number of investments in the model, i.e. the number of investments the investor has actually carried out. Kaplan and Strömberg (2004) find that time constraints are a significant concern to the investor.

Third, in the selection equation, each potential match has a valuation which increases with the quality of the startup as well as of the investor. Each of the two parties receive a fixed share (theta) of that value. This assumption is made as we cannot observe theta for every individual match and because it would complicate the model to an incomputable level. In the standard college admission problem on which this model is based, transfers would lead to multiple equilibria and the likelihood function might yield inconsistent results (Bresnahan and Reiss (1991)) and Hsu (2004) find only little evidence of transfers, with high reputation investors acquiring equity at a 10-14% discount. Transfers are therefore ruled out, and less promising startups cannot attract better investors by offering a higher share in the company.

Because the equilibrium matching we observe in the market is not just an aggregate of individual agents' choices, but of relative choices and interactions depending on all other agents in the market, the error terms must be integrated simultaneously for the maximum-likelihood estimation. With the current computing power available this is not feasible. I therefore apply the integration developed by Klein (2018), which, for the problem at hand, follows Sørensen (2007) and solves the estimation by Bayesian inference using a Gibbs sampling algorithm.

3 Data and variables

3.1 Database

The data used for the empirical analysis comes from the Crunchbase [2] database. The database was created in 2007 and is supplemented with data going back until 1990 but only few observations on funding rounds before 2000. The data is updated daily and partly crowd-sourced by updates from over 4000 venture partners that upload their portfolios, and over 600'000 executives, entrepreneurs and investors, supported by various algorithms screening newspapers and Crunchbase employees to supplement and control the quality of the data by hand. It contains information on 137'975 investors, 301'844 unique funding rounds, 459'317 investments, 100'999 acquisitions, 18'174 IPOs and 928'244 organisations. The dataset contains less observations for earlier years and started to grow in scale especially after 2010. As with most datasets on VC investments, the completeness of the data might be an issue. But especially for more recent years the coverage seems high and the broad spectrum of information on companies and investors over time that can be leveraged for this analysis is a big advantage. Aggregate statistics on VC funding by country and year are similar to the ones from other more established datasets (Dalle et al. (2017)), but there should still be caution when interpreting the results as the way the data is gathered might cause some selection bias that is difficult to control for.

To construct the investor network and extract the social capital indicators, the full dataset is used. I then restrict the observations to funding rounds that took place after 2000. I combine the various observations from the datasets on funding round level and restrict it to the first documented funding round for each company, keeping only investments indicated as Seed round or Series A to assure the first documented funding round is not only reported as such because earlier funding rounds were not reported in the data. I further only keep funding rounds that took place in the first 10 years after the company's founding date to restrict the data to young companies. As the focus lies on the investor's social capital, and there are oftentimes multiple investors in each funding round, I restrict the sample to one observation per funding round with investor details of only the lead investor. This also satisfies the conditions for the one-to-many matching model suggested in Sørensen (2007) and discussed above in the model's assumptions. For the few cases where there are multiple lead investors documented, I simplify by picking the one lead investor with the most experience. I further restrict the sample to only contain investments

[2] http://www.crunchbase.com/

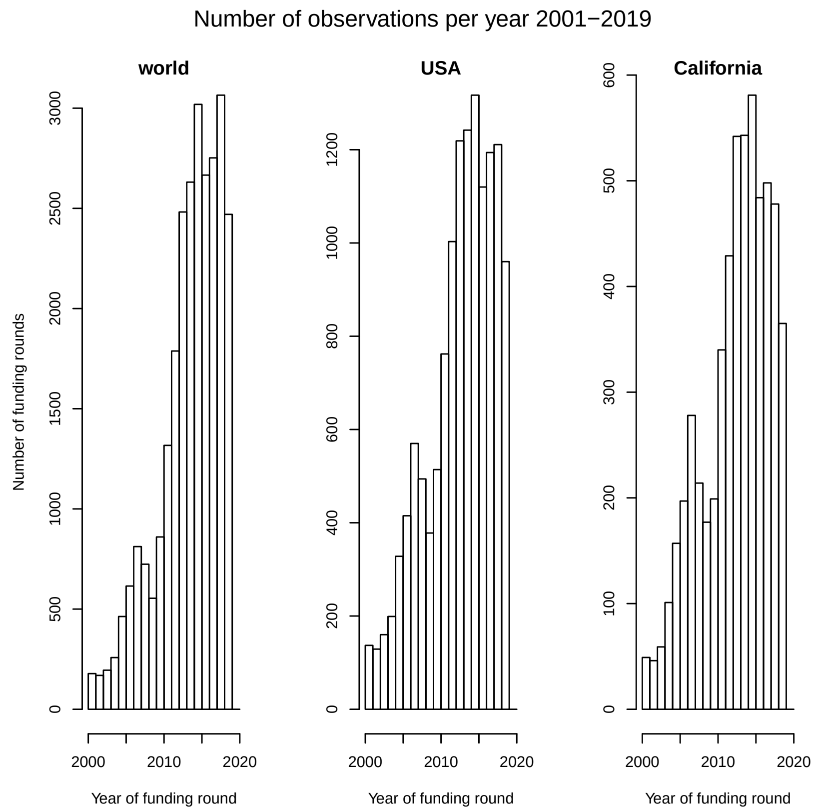

Figure 1: Number of observations between 2001 and 2019 by region

led by active investors, defined as investors having invested in at least 10 funding rounds in the respective market over the sample period. Furthermore, only funding rounds led by investor documented as venture capitalists and angel investors as well as a few undefined are kept. Given the strict exclusion conditions before, the undefined investors are assumed to be in the venture capital business. The cleaned dataset contains 27'018 first funding rounds worldwide with increasing observations towards the end of the dataset. Almost half of the businesses invested in are located in the USA (13'352) of which a large part is located in California (5'737). Figure 1 shows the similar relative evolution in the number of funding rounds over time for the different defined regions. Since data seems most complete for the California market and specially to comply with the requirements of the computationally intensive structural model, the main analysis is restricted to this specific market.

3.2 Dependent variables

I define four estimations with different dependent variables: The funding amount of the first funding round of the startup, the probability of the company having at least one subsequent funding round, the total subsequent funding received after the initial funding round, conditional on at least one additional funding round, and the company's exit probability through acquisition or exit. These last three variables represent measures for the success of the companies after they have received their first VC funding.

The initial funding amount enters into the model in logarithmic form because it is highly skewed to the right, with a few very high initial funding amounts. After controlling for matching in the structural model, the coefficient measures the impact of the variables on the willingness to pay or their ability to raise more funding from investment syndication, and might at least partly indicate their perceived risk which might be different depending on the investors characteristics and how well the investor is connected. The total subsequent funding amount enters in logarithmic form as well, and a positive coefficient in the structural model indicates the ability to attract further funding and overcome potential financial gaps, i.e. the valley of death. The probability of further funding rounds measures the ability to create value and confidence in the marketability of the product in the short run. The exit probability is widely used as a measure of success. It shows the success of the company in creating marketable value, but it is also the final goal of the VCI to accompany the investees to a successful exit and make a profit and therefore to a certain extent also represents the success of the investors. Exit is defined as either being acquired by another company or going public in an IPO.

3.3 Social capital

The question I try to answer in this empirical analysis is whether certain relational structures of investors in their network have an influence on their investment behaviour and on the companies they invest in. To do so, I construct a network of investors where each node represents an investor and any investor pair forms a tie when they co-invest in the same funding round. This eventually represents a one-mode representation of a two-mode network (investors and funding rounds as two types of nodes). Such investor pairs and their respective ties they formed can be seen as a potential channel for exchange of information, support and investment opportunities. This network is constructed from the full dataset, including not only VC investment rounds and VCIs, but all funding rounds and investors. When two investors invest in the same company together, this relationship cannot be assumed to last forever. I follow the approach taken by Alexy et al. (2012) and Hochberg et al. (2007) and assume a duration of 5 years with ties thus ending 5 years after the last co-investment. Similarly, I assume investors only to be active investors when they have a reported investment in the last 5 years. Figure 2 gives an overview of the network size over time.

Similar to the number of observations in the sample, the network size increases strongly over time. The average number of ties varies between 6.5 and 9. The resulting dynamic network is the basis to extract two measures for social capital, eigencentrality and constraint.

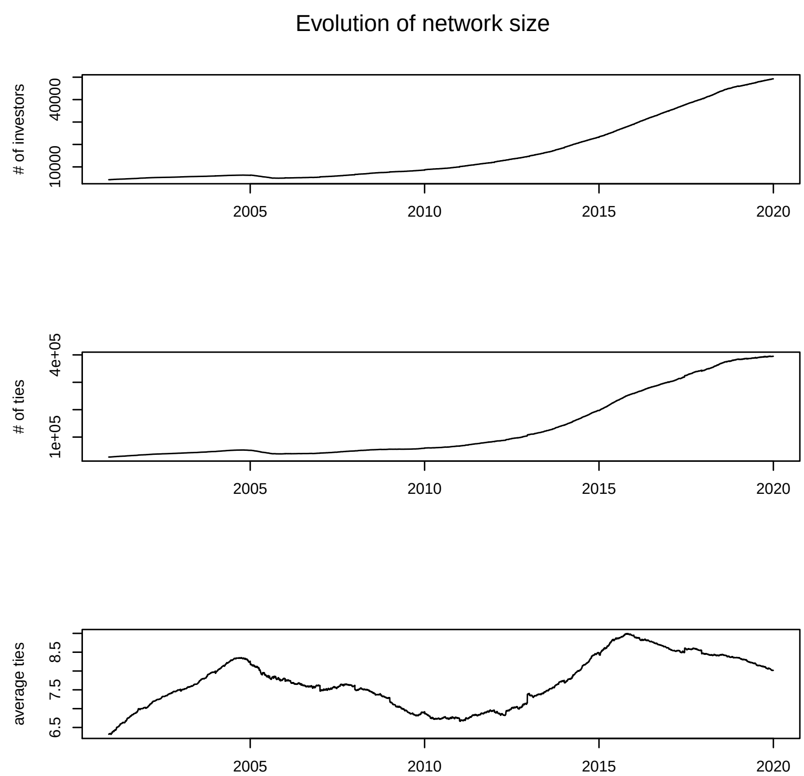

Figure 2: total investors, ties and average ties per investor in the unrestricted investor network

3.3.1 Eigencentrality

Every tie between two investors can be seen as a potential channel to exchange knowledge and can be leveraged to increase the investors investment possibilities and to gather resources for themselves or for the companies they have invested in. The more ties an investor has, the more potential exchange and support can happen. Furthermore, a tie to an investor that has many ties himself is likely to be more beneficial than a tie to an investor without any other connections. This is captured by the eigenvector centrality (eigencentrality) of the investor, the values of the first eigenvector of the graph adjacency matrix. This can be seen as arising from a reciprocal process, where an investor with high eigencentrality also increases the eigencentrality of his co-investors. High eigencentrality investors would usually be in big closely connected cliques of investors that are likely to operate in similar markets.

The eigencentrality measure is constructed for each investor at each funding round date. Since every tie has only a 5-year spell, the eigencentrality of each node represents the connections formed through co-investment over the 5 years prior to the funding round valued by the influence the tie partners have themselves. As the number of total investors in the network varies over time, and thus also the number of potential ties and the maximum possible eigencentrality, I take a scaled eigencentrality measure by dividing the eigencentrality of each investor by the maximum observed eigencentrality at the time of the funding round. This allows for better comparison over time and the scaled eigencentrality basically represents the relative influence of the investor at the time of investment.

3.3.2 Constraint

In network analysis, another important concept is that of structural holes. An illustrative graphical representation is shown in figure 3. When the only two-path connection between two nodes, say investor A and B, goes through a third node, investor C, that both A and B have a tie with, then investor C is in a structural hole. Or in terms of groups of investors, when the members of a closely connected group of investors that mainly co-invest with other investors from the same group, but less with investors from another group, then the investor (C) that co-invests at least once with a member of both groups is in a structural hole. Information flow from one group to the other has

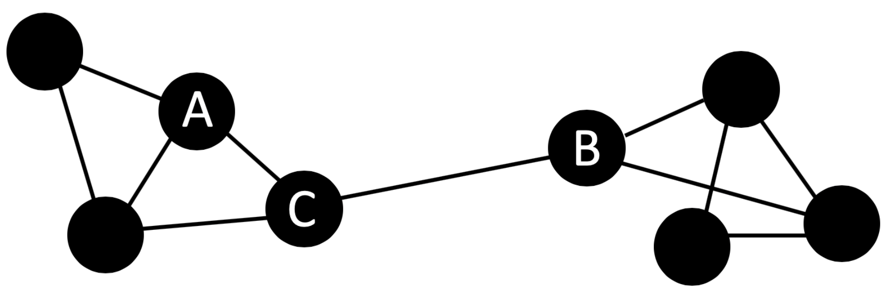

Figure 3: Illustration of structural hole

to go through investor C that occupies this structural hole and that investor can therefore act as a broker. The network constraint measures the extent to which a node's contacts are redundant or put differently it measures the absence of structural holes in a node's network of co-investments. The broker that has access to novel information from both groups is therefore less constrained, while a node for which all tie partners also form ties between each other has a higher constraint, as this node only gets information from within the closed group of "friends" and this same information is available to everyone in the group. In the investor network, a lower constraint might be leveraged by the investor to get access to high potential deals (sorting) and to novel information and unique networking possibilities that might be beneficial to his investees (influence). A lead investor with a low constraint might also be more likely to bring together investors with different target markets and different experiences to invest in the same funding round. To measure constraint in the investor network, I apply Burt's Constraint measure as defined in Burt (2004) as

$$C_i = \sum_j c_{i,j} = \sum_j \left[p_{i,j} + \sum_q p_{i,q}p_{qj}\right]^2, \quad i, j \neq q$$

where $c_{i,j}$ measures the dependence of investor i on investor j and $p_{i,j}$ measures the amount of network time and energy investor i puts directly into his relationship with investor j, which in this network with unvalued ties is simply the reciprocal of the total number of ties of investor i at the time of the funding round. $\sum_q p_{i,q}p_{qj}$ represents network time and energy investor i puts into his relationship with investor j indirectly through investor q.

In the further analysis and discussion, I will use the wording "higher social capital" to refer to investors that are less constrained and that have a higher eigencentrality. This wording is technically not entirely accurate and context specific. While a high eigencentrality can be associated to a higher value of social capital, a low constraint measure is rather descriptive and cannot be assigned to a high or low objective valuation. In theory a low constraint can be associated with opportunities, power or innovation but a high constraint can also be beneficial and is associated with a higher embeddedness and safety.

3.4 Clean energy

To identify a potential difference in effects for clean energy startups, a clean energy indicator is included in the model. This includes organisations indicated in the Crunchbase data as "clean energy", "renewable energy", "wind energy", "solar", "biofuel" and "biomass energy" as well as "energy storage". The clean energy term is included linearly and as interaction term with the constraint and eigencentrality measure respectively to analyse a different influence of the social capital measures for the clean energy sector.

3.5 Control variables

To differentiate the impact of the investor's social capital in the investor network from general investing experience and potential ability that is attributed to that, I control for the previous investment experience, measured as the total number of investments the investor has conducted prior to the funding round. Experience is therefore closely related to the investors social capital but is likely easier to observe by the startups than the social capital, which is a more abstract concept. It is therefore to some extent also a measure of the investor's reputation in the market. For an easier display of the estimation outputs, I divide the previous investments by 100. As it can be assumed that above a certain number, an additional unit of experience will have a smaller effect I also include the squared term.

To control for time trends, I include year fixed effects. The year fixed effects are especially important in the subsequent funding and exit estimations. Due to the very recent data, startups that got their first funding only recently had much less time to raise further money and much less time to build enough value to exit. The year fixed effects in the three success equations therefore not only capture year to year changes in the market environment but also controls for the censoring of the data, as future funding rounds and exits cannot be observed.

Also, the more time a startup takes until the first funding round, the more time it has to build value and it is likely that investors prefer older startups for that reason and because the longer a startup has already existed, to more information about it might be available and thus decrease the risk of moral hazard. I therefore include the logarithm of the age of the firm at funding round date in years. Furthermore, the variable serial entrepreneur captures the effects and the preferences of investors over startups founded by serial entrepreneurs, which is measured as the number of startups founded before the founding date of the company. I also include the investor count of the funding round linearly and in squared terms. More investors in a funding round can lead to more resources available to the firm. It might also be that bigger investment rounds require more investors for risk sharing. I further include two dummy variables, the investor type, which equals 1 if the lead investor is a single person and 0 if it is an organisation, and the investment type, which equals 1 when the first funding rounds is indicated as series A and 0 when it is a seed round.

3.6 Descriptive statistics

Descriptive statistics of the variables for the analysed California subsample are shown in table 1. Table 9 and 10 in the appendix show the statistics for the USA subsample and for the world data respectively.

Because I defined an investor as active in a respective market as having made more than 10 investments in this respective market, the California data is marginally more restricted and tends toward more active investors, but the patterns and differences between the markets are similar when comparing the markets including all investors. Only eigencentrality, constraint and experience show stronger differences, as the measure of previous investments

Table 1: **Descriptive Statistics California**

	mean	median	min	max	stdev
log fr raised amount	14.715	14.914	9.210	20.668	1.478
additional funding rounds	0.623	1.000	0.000	1.000	0.485
log subsequent funding	10.347	15.331	0.000	23.931	8.170
exit	0.261	0.000	0.000	1.000	0.439
eigencentrality	0.176	0.079	0.000	1.000	0.217
constraint	0.063	0.018	0.003	1.284	0.183
experience	2.233	1.040	0.000	21.040	2.996
clean energy	0.014	0.000	0.000	1.000	0.116
serieal entrepreneur	0.531	0.000	0.000	13.000	1.122
age at funding round (log)	0.296	0.394	-5.900	2.295	0.967
investor count	3.736	3.000	1.000	95.000	3.877
investment type series A	0.478	0.000	0.000	1.000	0.500
investor type person	0.038	0.000	0.000	1.000	0.192

by construction is correlated to these measures.

California shows a higher mean for all outcome variables compared to all of the USA and especially compared to the world average. Further, the social capital measures show higher values for California, respectively a lower constraint. This might be due to a more active venture capital market in California in general, but potentially also due to better coverage of Californian investments in the database. The average initial funding amount for California is US$ 2.46 million. Subsequent funding amount, which is the sum of multiple funding rounds, is one third lower than the initial funding amount as it also includes the companies without any further funding round and 62% of the companies raise another funding round. The mean conditional on receiving subsequent funding would be US$ 11.9 million. The exit rate with 26% is rather high, especially since this measure is a lower limit as newly funded companies might still exit in the next years which is not covered by the data. The average company gets first funding from investors after somewhat more than two years of existence and has 3.7 investors per initial funding round on average. Surprisingly, California has the lowest percentage of companies that were categorized as "clean energy" (1.4 %). Most investors are organisations (96%) and the data contains almost half-half series A and seed funding rounds.

Looking at the correlation in table 2, we see that the social capital measures as well as experience show the highest correlation. This is underscored by the correlation of 0.51 between eigencentrality and experience, which is not surprising since higher experience, measured as previous investments, leads to more ties to other investors and increases the eigencentrality of the investor. Having more ties also increases the chance of an investor being in a structural hole and thus being less constrained. Looking at multicollinearity, the variance inflation factor for a regression without interaction terms shown in the last column of table 2, I find some, but not concerning multicollinearity for eigencentrality and experience. With the full regression including squared and interaction terms with cleanenergy, the VIF for the interaction terms is still low, as expected from the low correlation cleanenergy shows with any of the other regressors, but the main and squared terms of eigencentrality and experience show VIF scores all around 10. This is slightly aggravated by the collinearity of eigencentrality and experience, but mainly it is the structural collinearity stemming from the correlation between the main and the squared terms. As we will see in the estimations, the potentially bigger standard errors do not seem to cause relevant problems for the inference.

Table 2: **correlations**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	VIF
(1) log fr raised amount														
(2) additional funding rounds	0.09													
(3) log subsequent funding	0.15	0.99												
(4) exit	0.1	0.13	0.13											
(5) eigencentrality	0.15	0.06	0.08	0.12										1.78
(6) constraint	-0.17	0.04	0.03	-0.01	-0.21									1.10
(7) experience	0.00	-0.09	-0.09	-0.11	0.51	-0.2								1.77
(8) clean energy	0.03	0.04	0.05	-0.01	0.01	0.00	-0.03							1.01
(9) serieal entrepreneur	0.03	0.02	0.03	0.00	0.05	-0.04	0.02	-0.01						1.04
(10) log age at funding round	0.22	-0.11	-0.1	-0.06	-0.07	-0.02	-0.02	0.03	-0.09					1.10
(11) investor count	0.13	0.02	0.03	-0.05	0.07	-0.1	0.07	-0.04	0.14	-0.03				1.10
(12) series a	0.6	0.11	0.16	0.22	0.12	-0.04	-0.06	0.04	-0.03	0.2	-0.16			1.40
(13) investor is person	-0.1	0.00	-0.01	-0.02	-0.12	0.05	-0.13	-0.02	0.02	-0.04	0.09	-0.14		1.06

4 Baseline estimation and sorting on observed characteristics

I now turn to the empirical baseline analysis without controlling for unobserved sorting in the market using OLS and probit estimates. While every firm is represented only once in the dataset, investors invest in multiple startups and one could suspect some serial correlation. I estimated the models with and without clustering at the investor level which lead to only minor changes in standard errors and no change in significance. I ran each model with three different specifications which gave an estimation of the sorting on observed investor and company characteristics. All estimations include year fixed effects for the date of the initial funding round, but the corresponding coefficients are not shown in the tables. Column (3) in each table shows the full baseline model that serves as comparison for the estimation of the structural model.

4.1 First funding round amount

The first model estimates the amount raised in the first funding round of the startup with OLS. The results are reported in table 3. I first estimate the regression with only the regressors of interest, the social capital measures and the interaction with the clean energy dummy, in column (1). Eigencentrality as well as a lower constraint show a positive and significant effect on the amount raised in the first funding round. The social capital stays significant over all three specifications but changes significantly in scale and including controls raises the explanatory power substantially from an R^2 of 0.11 to 0.56. When controlled for the different observed investor and company characteristics, the effect of constraint is cut in half. The same goes for experience, where the coefficient decreases strongly after controlling for observed characteristics, indicating an upward bias from sorting.

Interestingly, the coefficient for eigencentrality increases in column (3). Having a closer look and running different specifications showed that it only increases when controlling for the investment type, whether it is a series A or a seed round, otherwise the effect decreases in magnitude with the inclusion of the other controls. Since the coefficient of series A is positive, highly significant and the data shows that series A lead investors have a higher eigencentrality over the whole distribution, I would expect the coefficient of eigencentrality to decrease when controlling for the funding round type. As I will discuss in more detail in the analysis of the bias from sorting on unobserved characteristics and the underlying matching game, this opposite evolution of the effect of eigencentrality can be explained by the competition between investors with different target markets.

In all three specifications the clean energy main term does not show a significant difference and the F-test for the

Table 3: **1st funding round investment amount**

OLS regression of log first funding round amount in the California market. Year fixed effects are excluded in the table for better oversight. 'investor is person' = 1 if lead investor is a single person and 0 if an organisation. 'series A' = 1 when the first funding round is series A, 0 when seed round. Robust standard errors are in parentheses.

	log raised amount (in USD)		
	OLS	OLS	OLS
	(1)	(2)	(3)
eigencentrality	1.180*** (0.237)	1.066*** (0.292)	1.637*** (0.219)
eigencentrality^2	−0.943*** (0.261)	−1.023*** (0.283)	−1.314*** (0.216)
eigencentrality*cleanenergy	0.630 (0.529)	0.514 (0.501)	0.725* (0.438)
constraint	−1.081*** (0.133)	−0.918*** (0.132)	−0.452*** (0.102)
constraint*cleanenergy	1.227 (1.487)	1.139 (1.493)	0.211 (1.013)
cleanenergy	0.050 (0.215)	0.130 (0.215)	0.026 (0.153)
experience		0.170*** (0.021)	0.026* (0.015)
experience^2		−0.017*** (0.001)	−0.007*** (0.001)
investor is person			−0.082 (0.073)
series A			1.939*** (0.033)
serial entrepreneur			0.030** (0.013)
log age at funding round			0.113*** (0.017)
investor count			0.116*** (0.006)
investor count^2			−0.001*** (0.0002)
Constant	15.855*** (0.131)	15.809*** (0.133)	13.614*** (0.114)
Year dummies	Yes	Yes	Yes
Observations	5,737	5,737	5,737
R^2	0.110	0.145	0.516
Adjusted R^2	0.107	0.141	0.513
Residual Std. Error	1.397 (df = 5712)	1.369 (df = 5710)	1.031 (df = 5704)
F Statistic	29.532*** (df = 24; 5712)	37.347*** (df = 26; 5710)	189.804*** (df = 32; 5704)

Note: *p<0.1; **p<0.05; ***p<0.01

linear combination of the clean energy terms is insignificant. But the full model in column 3 shows a significantly higher coefficient at the 10% level for the clean energy and eigencentrality interaction, indicating higher initial funding amounts into clean energy firms by investors that are well connected with other central investors. The effect of constraint seems to be lower for clean energy startups, and the F-test of the combined effect shows a p-value of 0.81, indicating that constraint has no effect on the initial funding amount for clean energy firms.

The signs for the other control variables are all as expected, with a higher funding amount for later series A funding rounds, a slightly higher funding amount for serial entrepreneurs and also older firm at the funding round date are able to raise more money. The more investors take part in the funding round, the higher the funding, but this is likely to be endogenous with bigger funding rounds potentially attracting more investors to spread the higher risk.

4.2 Probability of subsequent funding

The second baseline model estimates the probability of the firm to raise money in at least one funding round following the initial round. The results are shown in table 4. As in the initial funding amount estimation, I run three specifications with no controls in column (1), adding experience in column (2) and the full specification with all controls in column (3). The reported coefficients are the marginal effects at the mean with the corresponding standard errors and significance levels. First funding round year dummies are included in all three specifications. Only the years 2018 and 2019 show significantly lower probabilities, as expected given that these firms had less time to raise additional funding and the years 2002 to 2006 leading up to the financial crisis have significantly higher probabilities of raising additional funding.

Eigencentrality shows a strong positive coefficient in the first two specifications, but when controlling for the different investor, funding round and firm characteristics, the main coefficient decreases in magnitude, as I would have expected also in the first funding round estimation. But again, when estimating the full specification without and then with the control for series A funding, I find an increasing eigencentrality coefficient as in the first regression, with the difference that in this case the upward bias from sorting on the other observed characteristics is bigger in magnitude than the downward bias from competition. Also, the magnitude of the experience term decreases moving to the full model and the coefficient of constraint increases with the inclusion of controls and eventually becomes significantly positive on the 10% level in specification 3, indicating sorting on observable characteristics. The positive effect of a more constrained lead VCI indicates that for raising additional funding, it might be marginally

Table 4: **Probability of a subsequent funding round**

Probability that at least one additional funding round occurs after the initial round in the California market. Year fixed effects are excluded in the table for better oversight. 'investor is person' = 1 if lead investor is a person and 0 if an organisation. 'series A' = 1 when the first funding round is series A, 0 when seed round. Marginal effects at the sample mean are reported. Robust standard errors are in parentheses.

	Probability of additional funding rounds		
	Probit	Probit	Probit
	(1)	(2)	(3)
eigencentrality	0.229*** (0.087)	0.249** (0.114)	0.183 (0.115)
eigencentrality^2	−0.247** (0.109)	−0.276** (0.122)	−0.226* (0.122)
eigencentrality*cleanenergy	−0.221 (0.282)	−0.235 (0.286)	−0.260 (0.284)
constraint	0.034 (0.037)	0.049 (0.038)	0.067* (0.038)
constraint*cleanenergy	−0.736*** (0.247)	−0.741*** (0.248)	−0.749*** (0.246)
cleanenergy	0.220*** (0.060)	0.224*** (0.060)	0.234*** (0.056)
experience		0.014 (0.009)	0.009 (0.009)
experience^2		−0.002** (0.001)	−0.001* (0.001)
investor is person			−0.019 (0.038)
series A			0.067*** (0.016)
serial entrepreneur			0.006 (0.006)
log age at funding round			−0.045*** (0.008)
investor count			0.011*** (0.003)
investor count^2			−0.0001** (0.00004)
Constant			
Year dummies	Yes	Yes	Yes
Observations	5,737	5,737	5,737
Log Likelihood	−3,386.104	−3,379.452	−3,345.033
Akaike Inf. Crit.	6,822.208	6,812.903	6,756.066

Note: *p<0.1; **p<0.05; ***p<0.01

beneficial to a firm to take on an investor that invests within a more closely connected network of co-investors. But the marginal effect for clean energy firms is significantly lower, indicating a benefit from having a less constrained lead investor in the first funding round.

Whether clean energy firm have a significantly higher probability of raising subsequent funding, as suggested by the main term, depends strongly on the values of the investor's social capital. For interpretation, two otherwise equal average firms with lead investors that have average characteristics (constraint of 0.063, eigencentrality of 0.175), the clean energy startup is predicted to have a 14.1% point higher probability of raising another funding round, taking all interaction effects into account. Especially a low constraint score for the lead investor seems influential - a clean energy startup with an average constrained lead investor compared to one with a lead investor that has a one standard deviation higher constraint score is 13.7% points more likely to raise additional funding. But again, these results are suggestive, as the significance changes with the specific values of social capital.

The experience of the investor is jointly significant on the 10% level for specification 3 with only small effects. When the first funding round was a series A round, it is more likely that there will be further funding rounds and also having more investors in the first funding round increases the probability of further rounds. The founders being serial entrepreneurs does not have a significant effect. Interestingly, the older the firm is at the initial funding round, the lower the chance of an additional funding round. Maybe older firms are more likely to already have a marketable product and an own income and therefore choose not to raise further funding or firms that take longer to the first funding round have less potential and therefore are less likely to get further funding.

4.3 Total subsequent funding

Conditional on a startup raising at least one more funding round, model 3, shown in table 5, estimates the effects on the total amount raised in all funding rounds that follow. This reduces the sample to 3'573 observations. Since all the independent variables are taken or constructed for the initial funding round, it is not surprising that the explanatory power of this model with an R^2 of the full model of 0.16 is much lower. All three specifications include year fixed effects, which shows a time trend for the last 3 years with decreasing values leading up to 2020 but not for the other years.

Again, there is clear indication for sorting on observed characteristics. While the magnitude of the constraint and experience coefficients decrease moving to the full specification in column (3), eigencentrality behaves in line

Table 5: **Total subsequent funding amount**

OLS regression of log total funding amount of all funding rounds following the initial funding round in the California market. Year fixed effects are excluded in the table for better oversight. 'Investor is person' = 1 if lead investor is a person and 0 if an organisation. 'series A' = 1 when the first funding round is series A, 0 when seed round. Robust standard errors in parentheses.

	log total subsequent funding amount (in USD)		
	OLS	OLS	OLS
	(1)	(2)	(3)
eigencentrality	2.053*** (0.376)	1.512*** (0.470)	1.682*** (0.446)
eigencentrality^2	−1.696*** (0.451)	−1.411*** (0.488)	−1.441*** (0.465)
eigencentrality*cleanenergy	1.532** (0.755)	1.394* (0.760)	1.665** (0.822)
constraint	−0.312 (0.199)	−0.215 (0.200)	0.055 (0.193)
constraint*cleanenergy	−4.232 (2.575)	−4.075 (2.588)	−4.019 (2.612)
cleanenergy	0.471 (0.308)	0.508* (0.307)	0.447 (0.325)
experience		0.155*** (0.042)	0.045 (0.039)
experience^2		−0.014*** (0.003)	−0.007** (0.003)
investor is person			−0.005 (0.179)
series A			1.210*** (0.068)
serial entrepreneur			0.126*** (0.025)
log age at funding round			−0.072** (0.031)
investor count			0.069*** (0.011)
investor count^2			−0.001** (0.0002)
Constant	17.080*** (0.182)	17.109*** (0.183)	15.850*** (0.197)
Year dummies	Yes	Yes	Yes
Observations	3,574	3,574	3,574
R^2	0.060	0.067	0.163
Adjusted R^2	0.054	0.060	0.155
Residual Std. Error	1.715 (df = 3549)	1.709 (df = 3547)	1.620 (df = 3541)
F Statistic	9.479*** (df = 24; 3549)	9.821*** (df = 26; 3547)	21.544*** (df = 32; 3541)

Note: *p<0.1; **p<0.05; ***p<0.01

with the previous two baseline estimations. To get a grasp of the extent of the effect of eigencentrality for an average investor, a one standard deviation increase leads to approximately 20% higher subsequent funding. The effect for clean energy startups, with a significant interaction coefficient in all specifications, is even bigger with an approximate 73% increase in subsequent funding (0.55 increase in log funding). A well-connected lead investor in the first funding round increases the predicted future funding substantially. On the other hand, the investors constraint is insignificant in all three specifications, as is the interaction term with clean energy and the joint linear hypothesis test. While having an effect on the probability to raise an additional funding round, there is no indication that investor constraint would affect the amount raised for clean energy startups. Also the F-test for significance of the combined effect of all clean energy terms is not significant, indicating that clean energy firms overall do not raise more or less additional funding.

Experience is significant for the main and squared terms in specification (2) and remains jointly significant in the full model but with much lower magnitude. The control variables again all have the same signs as before, but with serial entrepreneurs having a significant and substantial positive coefficient. Older startups at the first funding round seem to raise less subsequent funding and having more investors initially increases the subsequent amount raised.

4.4 Exit probability

The last baseline regressions estimate the effects on the probability of exit, which for VC investments eventually is the final goal and the point where at least a few startups in the portfolio have to get to for the investor to make a profit. It is also the estimation which caught most attention in previous literature. The results are presented in table 6. As with all other models, year fixed effects are included in all specifications which as expected show a consistent time trend over the last 10 years, which is a common duration of venture captial funds and thus the time a startup is often given to eventually exit.

Contrary to the findings of Sørensen (2007) I do not find a positive effect of the investor's experience on the exit probability. The effect is small and insignificant for specifications (2) and (3). An estimation with experience as the only regressor shows a significant experience coefficient, but the more control variables are included, the lower the effect gets and eventually turns insignificant as soon as I control for the investment type, even with the exclusion of the social capital regressors. These findings are therefore similar to the findings from Sørensen's

Table 6: **Probability of exit**

Exit probability through IPO or acquisition in the California market. Year fixed effects are excluded in the table for better oversight. 'investor is person' = 1 if lead investor is a person and 0 if organisation. 'series A' = 1 when the first funding round is series A, 0 when seed round. Marginal effects at the sample mean are reported. Robust standard errors are in parentheses.

	Probability of exit		
	Probit	Probit	Probit
	(1)	(2)	(3)
eigencentrality	0.126* (0.069)	0.169* (0.088)	0.163* (0.087)
eigencentrality^2	−0.110 (0.084)	−0.145 (0.092)	−0.135 (0.091)
eigencentrality*cleanenergy	−0.146 (0.214)	−0.149 (0.217)	−0.148 (0.226)
constraint	−0.127*** (0.036)	−0.123*** (0.036)	−0.106*** (0.035)
constraint*cleanenergy	0.966** (0.389)	0.969** (0.399)	0.965** (0.434)
cleanenergy	−0.108** (0.042)	−0.108** (0.043)	−0.105** (0.044)
experience		0.00005 (0.008)	−0.007 (0.007)
experience^2		−0.0004 (0.001)	0.0002 (0.001)
investor is person			−0.019 (0.029)
series A			0.074*** (0.013)
serial entrepreneur			0.002 (0.005)
log age at funding round			−0.018*** (0.006)
investor count			0.015*** (0.004)
investor count^2			−0.001*** (0.0002)
Constant			
Year dummies	Yes	Yes	Yes
Observations	5,737	5,737	5,737
Log Likelihood	−2,707.559	−2,706.459	−2,682.055
Akaike Inf. Crit.	5,465.117	5,466.918	5,430.111

Note: *p<0.1; **p<0.05; ***p<0.01

baseline probit model and there is evidence for sorting on observed variables when social capital is not accounted for. Similarly, for the investor's constraint, when moving to the full model, the effect becomes smaller, indicating that a lower constraint might help the VCI to match with a more promising startup which causes sorting on observable characteristics. Constraint is significant for all specifications, but the positive effect of a lower constraint is only observed for non-clean energy startups. For a clean energy startup, the effect of constraint is significantly higher, which in sum would change the sign and lead to a positive effect for more constrained investors. As with the coefficients of the interaction terms in the other estimations, they should be interpreted at most as suggestive since the number of clean energy firms in the sample is very low. This is aggravated here as from the initially already low number of clean energy observations, only 18 of them exit, an issue that I will come back in the final discussion of results.

Eigencentrality of an average lead investor on the other hand, although again showing a substantial positive effect that does not decrease when moving to the full model, is only significant on the 10% level. As before, I find the same evidence for a bias from sorting on observed characteristics for eigencentrality. The effect is also not significantly different for clean energy startups. Overall, for two similar average startups with average lead investors, the estimation provides suggestive evidence that a clean energy startup has a 7%-points lower probability of exit. This would be in line with the findings of other authors that suggested lower returns for cleantech companies (e.g. Gaddy et al. (2016)).

For the rest of the control variables, it stands out that serial entrepreneurship does not affect the exit probability and that older firms at the funding round date, similar to the findings of the subsequent funding estimations, are less likely to eventually exit.

5 Structural Model - sorting on unobserved characteristics

The four baseline estimations have shown the presence of biases of the social capital and experience coefficients arising from sorting on observable characteristics. I now turn to the empirical estimation of the structural model that attempts to control for the effects of sorting on unobserved characteristics. Since the estimation of the structural model is computationally very intensive, it is estimated using Gibbs sampling [3]. The outcome equation in all four models is defined as in the full baseline models (columns (3)) such that the magnitude of the coefficients can be compared. The selection equation is the same in all four models.

The latent investor-company match valuation is identified assuming competition in the investment market, but this competition can only be assumed when the startups and the VCIs are actually competing in the same market. Geographically, the data is already restricted to the Californian market, but there will not be competition between investment opportunities that happen in different years. I follow Sørensen (2007) and split the data into half year markets, that is, VCIs and startups in funding rounds that took place in the first half and those that took place in the second half of each year are assumed to compete in the same market. This leads to 38 different markets within which all VCIs that invested in that time period compete for the best possible match with startups that got funding in the same period, and vice-versa.

With the computational power available, the number of feasible iterations is limited and set to 600'000 iterations, where the first half is used as burn-in period. For the selection equation, when experience and eigencentrality enter the model in the selection equation also in squared terms, the MCMC procedure has problems of consistently converging within the feasible number of iterations. A possible explanation for this, besides the defined equations having a rather high number of coefficients to estimate, is that experience and eigencentrality show high collinearity with their respective squared terms. This might cause difficulties for the MCMC estimation to differentiate properly between the two effects. While a selection equation with higher order terms would be preferred, a simpler selection equation without squared terms is applied to approximate the effects. With the simpler selection equation, all models seem to have converged relatively quickly and only the last 300'000 iterations are used to compute the mean coefficients and the standard errors. Figures 4 and 5 in the appendix report sample convergence plots for the first funding round amount estimation, with the other three estimations showing a similar convergence behaviour. The estimates prove to be consistent up to only very minor differences when running the estimation multiple times.

[3] the estimation is implemented in R using the package "matchingMarkets", developed by Klein (2018)

5.1 Sorting

The estimations of the four model specifications are shown in table 7 and 8. The lower section of the tables presents the results of the selection equation. While the models converge consistently when run multiple times, there are differences between the model estimates for the selection equation because of the correlation between the error terms.[4] The covariance, which captures the correlation between the effects of unobserved characteristics that influence both the latent valuation as well as the dependent variable in the outcome equation, is highly significant for the first three estimations and the exit regression shows a p-value of 0.085. In line with Sørensen (2007), I find a significant effect of sorting on unobserved characteristics, but the effect seems stronger in the estimations of dependent variables that are more closely related to the first funding round. Intuitively, unobserved characteristics that affect the (latent) valuation of an investor-startup match are likely to affect the willingness to invest more in a startup since this is directly related. If an investor for example observes a highly skilled management team, the investor is probably willing to invest a higher amount. But it might be more difficult to evaluate the characteristics of the other party of the investment match in terms of the importance for long term success and it is also likely that at least some of the characteristics will change over time. A management team might be replaced, or an investor that seemed to be a good fit to the startup might not continue to support the startup and be replaced by another lead investor.

In all estimations, the investor's constraint clearly seems to matter in terms of the investor's ability to match with higher valued startups. It is unlikely that the startup can actually observe the investor's constraint, at most to the extent that it might be observable that an investor tends to invest in different industries or with a varying group of other investors and is therefore perceived as having a more diverse background. In that case, startups would prefer more versatile investors over investors that always invest with the same peers in similar industries. Another explanation could be that a lower constraint within the network of investors does not necessarily make the investor more preferred among firms, but simply leads to the investor having access to better investment opportunities. In any case, being less constrained increases the probability to match with a higher valued startup. For two average investors that both would want to invest in the same company, a one standard deviation higher constraint c.p.

[4] see Sørensen (2007), p. 2760, the distribution of α is $\mathcal{N}(-M_\alpha^{-1}N_\alpha, M_\alpha^{-1})$ with $M_\alpha = \Sigma_\alpha^{-1} + \sum_{m=1}^{N}\left[\sum_{ij\in M_m} W_{ij}W'_{ij} + \sum_{ij\in\mu_m} W_{ij}W'_{ij}\right]$ and $N_\alpha = -\Sigma_\alpha^{-1}\bar{\alpha} + \sum_{m=1}^{N}\left[\sum_{ij\in M_m} -W_{ij}V_{ij} + \sum_{ij\in\mu_m} \delta W_{ij}\left(Y^*_{ij} - X'_{ij}\beta - V_{ij}\delta\right)\right]$ and therefore depends on the estimated error term of the outcome equation.

Table 7: **First funding round amount and subsequent funding probability - structural model**

Structural model with selection equation controlling for sorting on unobserved characteristics. Selection equation estimating the latent match valuation and the two outcome equations estimating the first funding round amount and the probability of subsequent funding respectively. Estimate, StdErr and p.value is the mean, standard deviation and the respective p-value of the simulated posterior distribution of the estimates. mfx shows the marginal effects at mean in the binary case.

	first funding round amount				additional funding probability			
	mean	mfx	Std.Dev	p-value	mean	mfx	Std.Dev	p-value
OUTCOME								
intercept	12.0494		0.1429	0.000	-0.5131	-0.1938	0.2147	0.017
eigencentrality	2.0096		0.2345	0.000	0.8724	0.3294	0.3214	0.007
eigencentrality^2	-1.7678		0.2316	0.000	-0.9309	-0.3515	0.3365	0.006
eigencentrality*cleanenergy	0.5213		0.4250	0.220	-0.7212	-0.2724	0.6978	0.301
constraint	-0.2073		0.0953	0.030	0.2872	0.1085	0.1162	0.013
constraint*cleanenergy	0.3145		0.7050	0.656	-2.0947	-0.7910	1.3043	0.108
cleanenergy	0.0411		0.1527	0.788	0.7768	0.2934	0.2440	0.001
experience	0.0715		0.0154	0.000	0.0407	0.0154	0.0226	0.072
experience^2	-0.0070		0.0010	0.000	-0.0033	-0.0012	0.0015	0.031
investor is person	0.1386		0.0881	0.116	0.2594	0.0980	0.1220	0.033
series A	1.6468		0.0312	0.000	0.0709	0.0268	0.0468	0.129
serial entrepreneur	0.0235		0.0108	0.030	0.0155	0.0059	0.0168	0.354
log age at funding round	0.0946		0.0135	0.000	-0.1222	-0.0461	0.0205	0.000
investor count	0.0877		0.0047	0.000	0.0245	0.0092	0.0077	0.002
investor count^2	-0.0010		0.0001	0.000	-0.0001	-0.0000	0.0002	0.691
Covariance	1.1034		0.0381	0.000	0.3960		0.0602	0.000
Year dummies	Yes				Yes			
SELECTION								
(s) eigencentrality	0.5455	0.1539	0.1647	0.001	-0.0215	-0.0061	0.1429	0.880
(s) constraint	-0.4056	-0.1144	0.0804	0.000	-0.3746	-0.1057	0.1067	0.000
(s) experience	0.0496	0.0140	0.0085	0.000	0.0733	0.0207	0.0117	0.000
(s) log age at funding round	0.0115	0.0032	0.0113	0.307	0.0033	0.0009	0.0122	0.787
(s) serial entrepreneur	0.0249	0.0070	0.0096	0.010	-0.0029	-0.0008	0.0103	0.776
(s) cleanenergy	0.0309	0.0087	0.1014	0.760	0.0598	0.0169	0.1167	0.608
(s) investor is person	-0.5154	-0.1454	0.0663	0.000	-1.0216	-0.2882	0.1498	0.000
(s) series A	0.4823	0.1360	0.0434	0.000	0.6056	0.1708	0.0621	0.000
Observations		5737				5737		

Table 8: **total subsequent funding and exit probability - structural model**

Structural model with selection equation controlling for sorting on unobserved characteristics. Selection equation estimating the latent match valuation and the two outcome equations estimating the total subsequent funding and the probability of exit respectively. Mean, std and p-val are the mean, standard deviation and the respective p-value of the simulated posterior distribution of the estimates. mfx shows the marginal effects at mean in the binary case.

	subsequent funding				exit probability			
	mean	mfx	std	p-val	mean	mfx	std	p-val
OUTCOME								
intercept	14.4377		0.3125	0.000	-0.6895	-0.1885	0.2097	0.001
eigencentrality	2.2088		0.4408	0.000	0.6700	0.1831	0.3213	0.037
eigencentrality^2	-1.9228		0.4695	0.000	-0.5526	-0.1510	0.3323	0.096
eigencentrality*cleanenergy	1.6823		0.8590	0.050	-0.6199	-0.1695	0.7130	0.385
constraint	0.3890		0.1754	0.027	-0.3582	-0.0979	0.1209	0.003
constraint*cleanenergy	-1.6558		2.5701	0.519	3.0400	0.8310	1.6127	0.059
cleanenergy	0.2125		0.3142	0.499	-0.4407	-0.1205	0.2674	0.099
experience	0.0795		0.0364	0.029	-0.0207	-0.0057	0.0277	0.455
experience^2	-0.0072		0.0028	0.009	0.0006	0.0002	0.0022	0.784
investor is person	0.6316		0.2050	0.002	0.0018	0.0005	0.1141	0.987
series A	0.9861		0.0737	0.000	0.2396	0.0655	0.0504	0.000
serial entrepreneur	0.1135		0.0245	0.000	0.0069	0.0019	0.0181	0.703
log age at funding round	-0.0719		0.0293	0.014	-0.0660	-0.0180	0.0210	0.002
investor count	0.0636		0.0101	0.000	0.0543	0.0148	0.0150	0.000
investor count^2	-0.0005		0.0002	0.011	-0.0028	-0.0008	0.0009	0.002
Covariance	0.9346		0.1133	0.000	0.1022		0.0593	0.085
Year dummies	Yes				Yes			
SELECTION								
(s) eigencentrality	0.1330	0.0375	0.1907	0.486	-0.0173	-0.0049	0.1377	0.900
(s) constraint	-0.5259	-0.1484	0.1080	0.000	-0.4067	-0.1147	0.0917	0.000
(s) experience	0.0537	0.0151	0.0173	0.002	0.0743	0.0210	0.0106	0.000
(s) log age at funding round	-0.0025	-0.0007	0.0194	0.899	0.0042	0.0012	0.0128	0.745
(s) serial entrepreneur	0.0171	0.0048	0.0163	0.294	-0.0023	-0.0006	0.0106	0.829
(s) cleanenergy	0.7526	0.2123	0.6567	0.252	0.0687	0.0194	0.1204	0.568
(s) investor is person	-0.8591	-0.2423	0.1700	0.000	-0.9848	-0.2778	0.1324	0.000
(s) series A	0.6588	0.1858	0.0942	0.000	0.6170	0.1741	0.0616	0.000
Observations	3574				5737			

reduces the probability of the investor to match with that company by approximately 2 to 2.7%. The range of this estimate stems from the model specification, where different outcome equations produce different error terms and different correlations of the unobserved variables that then influence both the latent valuation as well as the dependent variable in the outcome equation.

Similarly, startups prefer more experienced investors and have a strong preference of an organisation over a single person as their investor. Investors on the other hand prefer to invest in series A funding over seed rounds, but there are no significant preferences whether a startup is in the clean energy sector and also the age of the startup does not seem to influence the investors investment decision. While these coefficient show a more or less similar picture across all four estimations and allow for an interpretation at least in sign and significance, the effects of serial entrepreneurship and eigencentrality show a big difference in the first funding round amount estimation and require some further explanation of the workings of the structural matching model. The higher the effect of the unobserved characteristics that affect both the matching in the selection as well as the dependent variable in the outcome equation, captured by the correlation between the error terms, the greater the distribution of coefficients of the outcome and the selection equation are affected by the error terms of the respective other equation. That is, the coefficients in the selection equation contain additional information from the outcome equation and are not solely dependent on the identifying restrictions imposed on the valuation draws. The selection equation in the first funding round amount estimation therefore contains more information from the outcome equation, which intuitively makes sense as the amount raised in the first funding round is likely to be a good approximation of how a company is valued at the funding round date and unobserved characteristics are likely to affect both the dependent variable in the outcome equation as well as the valuation of the match. Whether a company exits years after the matching occurs on the other hand is less informative for the initial valuation of the match. Based on this additional information from the outcome equation, Eigencentrality shows no effect on the matching in the three success estimations, but with the high correlation of the error terms in the first funding round estimation the effect is high in magnitude and strongly significant. Assuming the amount raised in the first funding round is a good approximation of the subjective valuation of the startup at the funding round date, startups prefer well connected lead investors with a high eigencentrality. But the unobserved characteristics that impact the valuation of the match and the startup are less informative for future success and when the additional information for the estimation of the selection equation is gained from the outcome equation for the three success measures, which eventually represent actual post-funding round valuation, successful startups are not more likely to choose lead investors with high eigencentrality.

5.2 Bias from sorting on unobserved characteristics

In any case, the correlation of the error terms captures the sorting on unobserved characteristics that influence the valuation as well as the respective dependent variable in the outcome equation and hence controls for the sorting bias in the outcome equation. The coefficients should at least be less biased and represent the influence the investor has on the startup's success more accurately.

Similar to the pattern observed when controlling for the characteristic of the investment type (seed or series A round), the magnitude of the coefficients of level and squared eigencentrality increases and remains significant in all four regressions after controlling for sorting on unobserved characteristics. I would expect a better connected VCI with ties to other more influential VCIs to match with startups that have beneficial characteristics and therefore controlling for these should reduce the estimated influence. The estimations suggest the opposite. A VCI with high eigencentrality is likely to be connected with other even more influential VCIs and a high eigencentrality might increase the investors chances to invest in more promising startups. Since this study focuses only on lead investors, it could also reduce the probability of the investor to be the lead investor in the funding round due to a higher competition for that lead position. This might explain the fact that there is no effect of eigencentrality observed in the selection equation in the three success estimations. But it would not explain the increase in the coefficient of the outcome equation. A possible reason for the increase is that a high eigencentrality potentially does make the investor more attractive to startups, but having ties to other influential investors also means that the investor's usual target market might be more competitive in the presence of investors with higher eigencentrality, which can cause an opposite bias from sorting. When the latter effect is bigger, the estimated influence increases after controlling for that competition, i.e. for sorting on observed and unobserved characteristics. Compared to the baseline estimates, the effect of a one standard deviation increase in eigencentrality for an average investor increases by 12%, 137%, 29% and 13% in the four estimations respectively.

The increase in the eigencentrality coefficient in the baseline estimates when controlling for the series A investment type dummy might be caused by a similar process. On average, eigencentrality of series A lead investors is higher in the data. This means that controlling for series A might cause two opposite effects. One is that initial series A funded companies perform significantly better in all estimations, and sorting on the investment type should hence control for an upward bias caused by this. The other is that due to the higher competition in the investment market of series A funding rounds, indicated by a higher average eigencentrality, including the series A dummy might at least partly also control for the competition on these preferred funding rounds, which controls for a potential downward bias. In the baseline estimation the latter effect seemed to be stronger.

The effect of constraint on the other hand evolves as expected and in line with the findings in the baseline estimations when controlling for observed characteristics. After controlling for sorting on unobserved characteristics, the coefficient further increases (remember to think about constraint in opposite signs). Controlling for sorting on observed characteristics in the baseline model estimating the first funding round amount more than halved the effect of constraint. Controlling for sorting on unobserved characteristics in the structural model reduced the effect again by more than half to 46% the size of the full baseline model. For the exit probability estimation, where the sorting on observed characteristics reduced the magnitude of the coefficient by 17%, controlling for sorting on unobserved characteristics reduced it by another 8% compared to the full baseline estimation. The subsequent funding probability and total amount estimations are interesting cases where the increase in the constraint coefficient leads to significantly positive coefficients, suggesting that the influence of a lower constraint actually reduces the probability and the total raised amount of subsequent funding rounds. Put differently, startups with lead VCIs that are more strongly embedded withing a close group of co-investors are more likely to raise additional funding.

Experience is significantly positive in the selection equation and thus indicates that more experienced investors are matched in higher valued deals and I would expect an initial upward bias in the baseline estimation that the structural model corrects for. But contrary to the findings of Sørensen (2007), the coefficient of experience increases after controlling for sorting on unobserved characteristics in all four estimations. I cannot find an intuitive explanation for this positive change. Rather it might be caused by the correlation the variable shows with the social capital measures and especially with eigencentrality. Experience eventually shows a significantly positive but decreasing effect for all three funding estimations but keeps indicating no significant effect on the probability of exit.

Turning to the clean energy startups, I find no significant investor preference for clean energy firms. The sign and significance in the outcome equation also remain unchanged compared to the full baseline estimations. Only in the subsequent funding regression clean energy startup seem to be valued higher (but not significant), which leads to a lower corrected main coefficient in the outcome equation. This is due to the higher total subsequent funding for clean energy firms observed in the data. But as in the baseline model, the effect is insignificant. For the first funding round amount estimation, the interaction with eigencentrality decreases from 0.73 to 0.52 and becomes insignificant and the interaction with the investor's constraint remains insignificant. Looking at the influence of social capital on the probability of additional funding for clean energy firms, the interaction terms remain largely unchanged, but the p-value of the interaction of constraint and clean energy firms increases to 0.108. The stronger influence of eigencentrality on the total amount raised in subsequent funding rounds for clean energy startups still has the same magnitude and significance. For the effects on exit rates for clean energy companies, the magnitude of the coefficients decreases slightly for all three terms compared to the baseline probit and the results suggest a lower influence of social capital on the exit probability for clean energy firms, but with only the constraint interaction being significant on the 10% level.

For the remaining control variables, the estimates for the firms age at funding round, the funding round investor count and for serial entrepreneurship remain largely unchanged in all estimations. In line with the preference of investors of series A over seed funding, the effect of the first funding round being series A declines in all four estimations but remains significantly positive except for the probability estimation of an additional funding round. Most interestingly, startups seem to have a strong preference for an organisational over a single person investor, but after controlling for the sorting in the market, the influence of non-organisational investors significantly increases the probability and the total amount of subsequent funding rounds. This suggests that either organisations have generally better access to good deals, or that the startups' preferences for organisations over single person lead investors, given two otherwise identical investors, might be based on false assumptions.

5.3 Influence

After controlling for sorting, the results from table 7 and 8 now represent at least a less biased estimation of the actual influence the VCIs' social capital has on the startups success. I find that the investors eigencentrality has a significantly positive but decreasing influence on the amount of funding raised in the first funding round, on the additional funding probability and total amount, as well as on the final exit probability. The higher the eigencentrality, the lower the marginal effect. The magnitude of the effect is substantial. For an average startup-investor match, a one standard deviation increase in eigencentrality of the initial lead investor leads to a approximate increase in initial and total subsequent funding of 21% and 24% and to a 2.8 and 2.1%-point increase in the probability of the startup to raise at least one more funding round and to go public or be acquired, respectively. The investors constraint has a significant influence on all four measures, but with a different sign when it comes to additional funding probability and amount. While a less constrained lead VCI has a positive influence on the initial funding amount and the probability of exit, in terms of the probability and total amount raised in subsequent funding rounds, it is beneficial to the firm to have a first round lead investor that is strongly embedded within a more closed group of co-investors. But since less constrained investors are able to raise more funding in the first funding round and have a positive influence on the eventual exit probability of the startup, it is possible that the cause of the lower subsequent funding is that there is less need for additional funding for their investees, which on average also already have a higher initial valuation. This relationship would need further investigation in future studies.

Looking at the investors experience, it shows that more experienced investors are willing to invest more in the initial funding round and also have a positive influence of subsequent funding. This might be partly caused by more experienced investors having higher financial means in general. But contrary to the findings of previous literature, I do not find any evidence for a positive influence of experience on the exit probability. My results suggest that it is not the direct influence of experience, measured as number of previous investments, but rather the more beneficial contacts within the network of co-investments usually coming along with more experience, that cause the positive influence. Nevertheless, when a startup has the choice between different investor to take on, it might be difficult to observe the investors eigencentrality or constraint and going with the more experienced investor is easier to observe as a proxy for a well-connected investor.

After controlling for sorting, I do not find any evidence that investors social capital has a different effect on the initial funding amount for clean energy companies. But overall, the data provides suggestive evidence that clean energy startups are more likely to raise further funding and that an initial lead investor with a high social capital might be more important for such firms to raiser higher future funding amounts. A possible explanation for that is what has been stated in previous research, that the energy sector is very capital intensive in general and therefore higher funding is needed to achieve the necessary scaling of the technology. This would be in line with the findings in the exit regression that suggests a lower exit rate for clean energy startups even though total funding raised is higher. For an average firm and investor match, c.p. a clean energy firm is predicted to be 9.7% less likely to exit. This is also in line with previous research (Gaddy et al. (2016)) which argued that the return on investment might have been one of the reasons for the strong decline in clean energy investments after 2010. But the overall picture of the differences for clean energy firms is not very clear, the share of clean energy firms in the dataset is small and these results should therefore be taken as suggestive at most.

6 Discussion of results

This brings us to a discussion of the results and their limitations. The results show that a simple estimation of the effects of the investors' eigencentrality and constraint include a strong bias arising from sorting on observed as well as unobserved characteristics in the investment market. Controlling for observed characteristics reduces that bias. Some of the potential unobserved characteristics the structural model controls for, like the number of patents or the funders' education, might be observed with the right data at hand and therefore included in the baseline models. This could eventually improve the baseline estimates and reduce the observed bias from sorting on unobserved characteristics in the structural model. Nevertheless, the outputs from the structural model show a significant correlation between the error terms of the selection and the outcome equation. This sorting on unobserved characteristics leads to significant differences in magnitude of the social capital coefficients, suggesting that for an analysis of the influence of the VCIs' social capital on the success of their investees, it is important to be aware of the sorting that occurs between investors.

The results from the selection equation suggest that it is beneficial for investors to co-invest with a more diverse set of investors, as this improves their chances to be the lead investor in a funding round of more promising startups. This notion is supported by my results throughout all estimations, suggesting a very strong downward bias of the constraint coefficient. Even though for subsequent funding a low constraint might be disadvantageous, it increases the probability of an eventual exit of the startup, which is the final goal of both the VCI as well as the startup. Additionally, from the startups' perspective, a less constrained investor also invests significantly more in the first funding round. The magnitude of the effect of constraint on the amount raised in the first funding round I find in the empirical analysis is one third lower than the findings of Alexy et al. (2012) that do not control for the sorting on unobserved characteristics but include different covariates. These results provide suggestive evidence for the hypothesis that a more diversified and well-connected investor has a lower perceived risk because he has better information on the market environment and is therefore willing to invest more initially. But as discussed in the methodology, one of the assumptions that has to be made to identify the model is a fixed sharing rule. An alternative explanation to the positive effect of high social capital on the first funding amount could therefore be that less constrained investors tend to buy in with a higher percentage of ownership of the startups and therefore pay more.

A very interesting and unexpected result is the increasing estimated effect of the investors eigencentrality when controlling for sorting on funding round type and unobserved characteristics and I do not find clear evidence that a high eigencentrality improves the VCIs' chances of matching with higher valued startups. A high eigencentrality of the lead VCIs arises when they continually co-invest with other very active and likely more successful investors. A potential underlying mechanism to explain this is that the more successful and influential investors eventually become, the better the deals they strive for, and thus the higher the probability of having to compete with other more influential investors. Nevertheless, especially from the startup perspective, I find that a lead investor in the first funding round with high eigencentrality has a positive influence on the firms funding and on the exit probability. These results suggest that the non-financial resources provided by better connected VCIs' bring more benefits to the firm.

Trying to analyse the differences in the effects of social capital for cleantech startups turns out to be very difficult. For once, there are already very few clean energy startups in the sample and the necessary restriction that had to be made in order to estimate the models further reduces this number to 78 cleantech startups. Although the attrition rate from the restrictions put on the sample was similar to the non-cleantech firms, the resulting number of observations is low and there is a significant risk of selection bias, in terms of only the most promising clean energy firms making it in the sample. Furthermore, my results do not show a clear trend of differences over all estimations. A significant difference in effects is only observed for eigencentrality on the total amount of subsequent funding, conditional on at least one additional funding round having occurred, and for constraint on the exit probability. There is some suggestive evidence that a higher social capital of the lead investor increases the total subsequent funding amount raised, which could help to alleviate the higher capital requirements to scale clean energy technologies. But then again, there is no evidence that this increases the eventual exit probability of the cleantech startup.

Overall, I find a strong and significant influence of the lead VCI's social capital on the startups ability to raise funding and on eventual exit, but no clear difference in effects for clean energy firms. These results should not be generalised. The structural model by Sørensen (2007), which is also very computationally expensive, required additional restriction for the chosen subsample. Thus, the results are only for VC investments of active investors in the Californian investment market. These observations are likely to be above average in terms of startup success probabilities and investor success and reputation. As with most data on VC investments, there also remains some doubts about the completeness of the initial data and a potential additional bias towards more successful startups that are potentially overrepresented.

Furthermore, there is a possibility that some of the results are driven by VCIs which are the lead investor in many different funding rounds. While clustering of standard errors in the baseline estimations did not change the results substantially and therefore a simple error structure also for the structural model was assumed, it might be that the estimated coefficients are mainly driven by a few highly active and therefore very influential investors in the data. I therefore also run the baseline estimations with the funding rounds of the four most active investors excluded. This reduces the sample by 555 observations (10%) which causes minor changes for most estimates. Only the clean energy interactions and the eigencentrality change significantly. The bigger changes in the clean energy interaction terms show the sensitivity of these results to the chosen sample and the necessary caution with the respective interpretation. The eigencentrality on the other hand increases strongly in magnitude in all four estimations, supporting the drawn conclusions of a strong positive effect of eigencentrality, but raising caution when interpreting the magnitude of the coefficient.

7 Conclusion

Facing a changing climate, new innovation especially in clean energy technologies are needed and venture capital has been a prominent instrument to help bringing innovative technologies to the market. This work contributes to the understanding of the enabling role venture capital investors can play in that process. More specifically, I define the venture capitalist's social capital as the investors eigencentrality and Burt's constraint in the investor network of investment syndication, and analyse the influence the social capital has on the startups first round funding amount, subsequent funding probability and total funding amount and the eventual probability of acquisition or IPO. I add to the literature by differentiating between the effects from sorting in the market and the actual influence from social capital, which allows to take on a firm and not an investor perspective in terms of added value to the startup's success. I do so by using a structural model introduced by Sørensen (2007) and empirically analyse investment data for the Californian market from 2001 to 2019. While the focus lies on the implications for cleantech, I could not identify a differentiated effect for cleantech firms. But I find that a lead investor in the first venture capital funding round that is better connected and is more influential, i.e. has a higher eigencentrality, has a significantly positive influence on the amount of funding in the first and in subsequent funding rounds and increases the probability of an eventual exit of the company. Furthermore, I find that it is beneficial for investors to co-invest with a more diverse set of investors, as this improves their chances to be the lead investor in a funding round of more promising startups and while a low constraint of the lead investor also increases the initial funding amount and the probability of exit, a highly constrained lead investor might be beneficial to raise more subsequent funding.

These results for the Californian market of active venture capital investors and startups from various sectors suggest, that well-connected investors positively influence the startups' success and therefore increase the chance of innovative technologies to reach marketability. The better connected to other influential investors and the less constrained to resources from a close group of co-investors the venture capital lead investor is, the higher the chance of success of the innovative startup. While in the data investors do not seem to have any preference of cleantech over non-cleantech startups, having better networked investors to invest into promising firms in the sector could benefit the scalability of clean energy innovation. For potential future research, it would be interesting to evaluate whether an increase in subsidies for startups or sectors attracts more successful and better-connected investors. If it does, then the positive influence of the venture capitalists' social capital could help to leverage the effect of these subsidies. Further possible research that could not be covered in this study is to look into different measures of social capital, whether other measures yield similar conclusions and if potential measures constructed from a two-mode network (keeping the level of the funding round) would alter the results. Other areas of potential future research could include expanding the analysis to different regions and also conducting a more sector specific analysis. Additionally, the results in this research have shown that eigencentrality is negatively biased when not controlled for the underlying competition in the investment market. While intuitively this could be explained by more competitive investment markets of high eigencentrality investors, it might be interesting from an investor's perspective to analyse this process in more detail. Yet foremost, since I was not able to find conclusive differences in effects for clean energy firms from the accessed data, it would be interesting to see in future research if there actually is a different effect of social capital in the clean energy sector.

References

Alexy, Oliver T., Joern H. Block, Philipp Sandner, and Anne L. J. Ter Wal, "Social capital of venture capitalists and start-up funding," *Small Business Economics*, November 2012, *39* (4), 835–851.

Bresnahan, Timothy F. and Peter C. Reiss, "Empirical models of discrete games," *Journal of Econometrics*, April 1991, *48* (1-2), 57–81.

Burt, Ronald S., "Structural Holes and Good Ideas," *American Journal of Sociology*, September 2004, *110* (2), 349–399.

_ , *Brokerage & Closure*, New York, NY: Oxford University Press, 2005.

Chemmanur, Thomas J., Karthik Krishnan, and Debarshi K. Nandy, "How Does Venture Capital Financing Improve Efficiency in Private Firms? A Look Beneath the Surface," *The Review of Financial Studies*, December 2011, *24* (12), 4037–4090.

Clercq, Dirk De, Vance H. Fried, Oskari Lehtonen, and Harry J. Sapienza, "An Entrepreneur's Guide to the Venture Capital Galaxy," *Academy of Management Perspectives*, August 2006, *20* (3), 90–112.

Dalle, Jean-Michel, Matthijs den Besten, and Carlo Menon, "Using Crunchbase for economic and managerial research," OECD Science, Technology and Industry Working Papers 2017/08 November 2017.

Faria, Ana Paula and Natália Barbosa, "Does venture capital really foster innovation?," *Economics Letters*, February 2014, *122* (2), 129–131.

Gaddy, Benjamin, Varun Sivaram, and Francis O'Sullivan, "Venture Capital and Cleantech: The Wrong Model for Clean Energy Innovation," MIT Energy Initiative Working Paper July 2016.

Gale, D. and L. S. Shapley, "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, January 1962, *69* (1), 9–15.

Geronikolaou, George and George Papachristou, "Venture Capital and Innovation in Europe," *Modern Economy*, 2012, *03* (04), 454–459.

Gornall, Will and Ilya A. Strebulaev, "The Economic Impact of Venture Capital: Evidence from Public Companies," *SSRN Electronic Journal*, November 2015.

Hirukawa, Masayuki and Masako Ueda, "Venture Capital and Innovation: Which is first?," *Pacific Economic Review*, October 2011, *16* (4), 421–465.

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu, "Whom You Know Matters: Venture Capital Networks and Investment Performance," *The Journal of Finance*, February 2007, *62* (1), 251–301.

Hsu, David H., "What Do Entrepreneurs Pay for Venture Capital Affiliation?," *The Journal of Finance*, August 2004, *59* (4), 1805–1844.

_ , "Venture Capitalists and Cooperative Start-up Commercialization Strategy," *Management Science*, February 2006, *52* (2), 204–219.

International Energy Agency, "Special Report on Clean Energy Innovation - Accelerating technology progress for a sustainable future," Technical Report 2020.

IPCC, "Summary for Policymakers. In: Global Warming of 1.5 C. An IPCC Special Report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty," Technical Report 2018. [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)].

Kaplan, Steven N. and Per Strömberg, "Characteristics, Contracts, and Actions: Evidence from Venture Capitalist Analyses," *The Journal of Finance*, October 2004, *59* (5), 2177–2210.

Klein, Thilo, "Analysis of Stable Matchings in R: Package matchingMarkets," *The Comprehensive R Archive Network*, 2018.

Kortum, Samuel and Josh Lerner, "Assessing the Contribution of Venture Capital to Innovation," *The RAND Journal of Economics*, 2000, *31* (4), 674–692.

Lerner, Joshua, "The Syndication of Venture Capital Investments," *Financial Management*, 1994, *23* (3), 16–27.

Nanda, Ramana, Ken Younge, and Lee Fleming, "Innovation and Entrepreneurship in Renewable Energy," in "Innovation and Entrepreneurship in Renewable Energy," University of Chicago Press, July 2014, pp. 199–232.

Paglia, John K. and Maretno A. Harjoto, "The effects of private equity and venture capital on sales and employment growth in small and medium-sized businesses," *Journal of Banking & Finance*, October 2014, *47*, 177–197.

Popp, David, Jacquelyn Pless, Ivan Hascic, and Nick Johnstone, "Innovation and Entrepreneurship in the Energy Sector," in "The Role of Innovation and Entrepreneurship in Economic Growth," University of Chicago Press, January 2020.

Puri, Manju and Rebecca Zarutskie, "On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms," *The Journal of Finance*, December 2012, *67* (6), 2247–2293.

Sapienza, Harry J., Sophie Manigart, and Wim Vermeir, "Venture capitalist governance and value added in four countries," *Journal of Business Venturing*, February 1996, *11* (6), 439–469.

Sorenson, Olav and Toby Stuart Toby E., "Syndication Networks and the Spatial Distribution of Venture Capital Investments," *American Journal of Sociology*, May 2001, *106* (6), 1546–1588.

Sørensen, Morten, "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital," *The Journal of Finance*, December 2007, *62* (6), 2725–2762.

8 Appendix

Table 9: **Descriptive statistics USA**

	mean	median	min	max	stdev
log fr raised amount	14.473	14.732	6.908	20.668	1.628
additional funding rounds	0.609	1.000	0.000	1.000	0.488
log subsequent funding	9.926	14.691	0.000	23.931	8.076
exit	0.244	0.000	0.000	1.000	0.429
eigencentrality	0.123	0.030	0.000	1.000	0.193
constraint	0.115	0.027	0.003	1.284	0.261
experience	1.632	0.610	0.000	21.040	2.570
clean energy	0.017	0.000	0.000	1.000	0.128
serieal entrepreneur	0.440	0.000	0.000	23.000	1.079
age at funding round (log)	0.384	0.494	-5.900	2.300	1.020
investor count	3.271	2.000	1.000	95.000	3.254
investment type series A	0.458	0.000	0.000	1.000	0.498
investor type person	0.036	0.000	0.000	1.000	0.186

Table 10: **Descriptive Statistics World**

	mean	median	min	max	stdev
log fr raised amount	14.158	14.376	6.908	21.454	1.703
additional funding rounds	0.533	1.000	0.000	1.000	0.499
log subsequent funding	8.500	12.612	0.000	23.931	8.087
exit	0.172	0.000	0.000	1.000	0.378
eigencentrality	0.076	0.006	0.000	1.000	0.160
constraint	0.192	0.049	0.002	1.284	0.338
experience	1.329	0.430	0.000	22.120	2.366
clean energy	0.017	0.000	0.000	1.000	0.128
serieal entrepreneur	0.319	0.000	0.000	24.000	0.944
age at funding round (log)	0.438	0.560	-5.900	2.300	1.032
investor count	2.693	2.000	1.000	95.000	2.674
investment type series A	0.402	0.000	0.000	1.000	0.490
investor type person	0.031	0.000	0.000	1.000	0.173

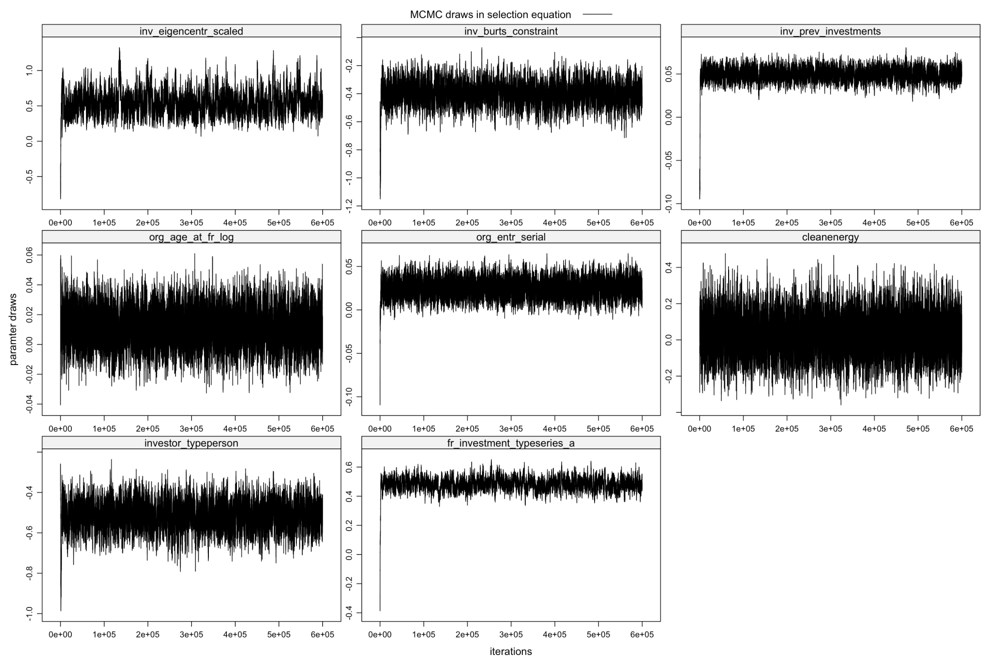

Figure 4: MCMC draws of the coefficients in the selection equation of the first funding round amount estimation

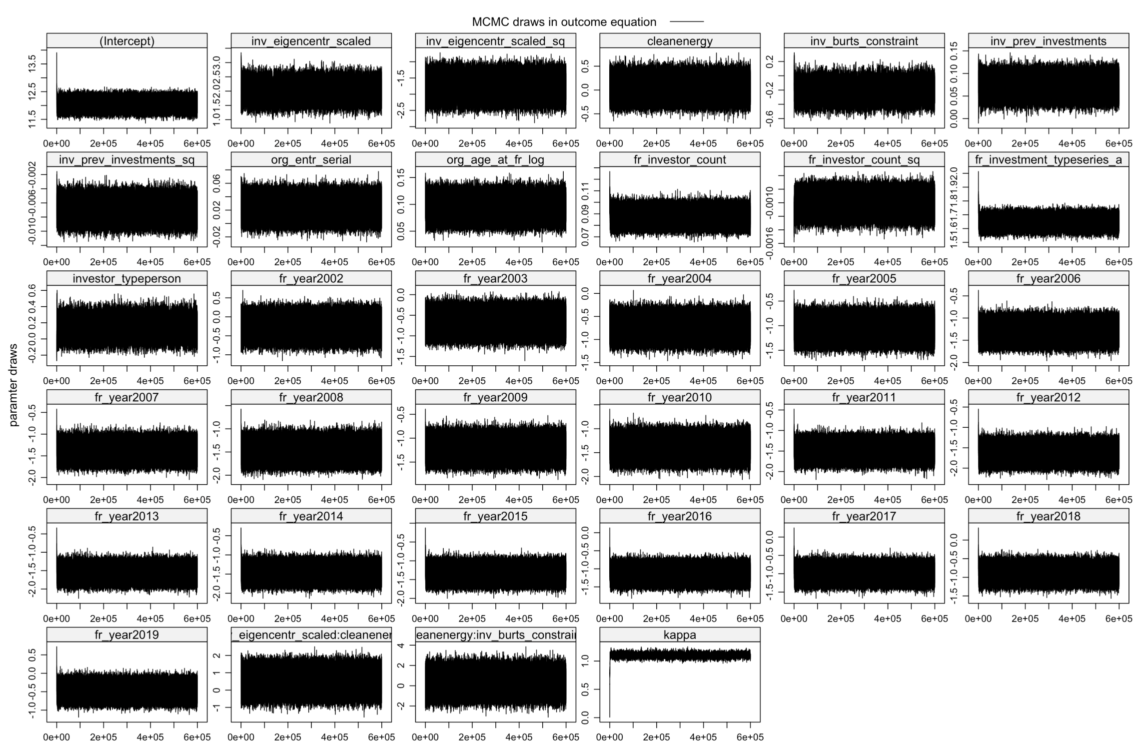

Figure 5: MCMC draws of the coefficients in the outcome equation of the first funding round amount estimation